Author's accepted manuscript: This is the peer-reviewed version of the following chapter: Terbeck, S., & Francis, K. B., (2018). *We should if we could, but we cant. Experimental problems with moral enhancement*. In: Hauskeller, M. (Ed.), *Moral Enhancement*. Cambridge: Cambridge University Press.

The Experimental Psychology of Moral Enhancement: We Should If We Could, But We Can't

Sylvia Terbeck* and Kathryn B. Francis

*Sylvia.terbeck@plymouth.ac.uk, University of Plymouth

Abstract

In this chapter we will review experimental evidence related to pharmacological moral enhancement. Firstly, we will present our recent study in which we found that a drug called propranolol could change moral judgements. Further research, which also investigated this, found similar results. Secondly, we will discuss the limitations of such approaches, when it comes to the idea of general "human enhancement". Whilst promising effects on certain moral concepts might be beneficial to the development of theoretical moral psychology, enhancement of human moral behaviour in general – to our current understanding – has more side-effects than intended effects, making it potentially harmful. We give an overview of misconceptions when taking experimental findings beyond the laboratory and discuss the problems and solutions associated with the psychological assessment of moral behaviour. Indeed, how is morality "measured" in psychology, and are those measures reliable?

1. Experimental Studies on Psychopharmacology and Human Morality

Recently, studies have begun to elucidate the neural basis of human moral behaviour, including neural correlates of moral action and decision making.[1] Initial studies used fMRI (functional magnetic resonance imaging) to determine areas of the brain associated with moral reasoning. In their widely-cited study, Greene, Sommerville, Nystrom, Darley, and Cohen used fMRI to determine the underlying neural mechanisms of moral judgements.[2] The study involved participants reading vignettes about moral dilemmas, such as the footbridge and the switch dilemma.[3] The authors categorized dilemmas as either personal or impersonal based on features of the scenario (i.e., up-close versus distant, involving redirected versus direct harm, etc.). They found that personal dilemmas were more strongly associated with activations in brain regions involved in emotional processing (such as the ventromedial prefrontal cortex) whereas impersonal dilemmas activated areas of working memory (such as the dorsolateral prefrontal cortex). In addition, the type of judgement individuals made also

1 S. Terbeck, J. Savulescu, L. P. Chesterman, and P. J. Cowen, 'Noradrenaline Effects on Social Behaviour, Intergroup Relations, and Moral Decisions', *Neurosci Biobehav Rev.* 66 (2016), 54–60.

2 J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen, 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science* 293:5537 (2001), 2105–2108.

3 P. Foot, *Virtues and Vices and Other Essays in Moral Philosophy* (New York, USA: Oxford University Press, 1978). See also J. J. Thomson, 'Killing, Letting Die, and the Trolley Problem', *Monist* 59:2 (1976), 204–217.

led to differences in brain activation patterns. Specifically, when making a utilitarian judgement (compared to a deontological judgement) in personal (but not impersonal) moral dilemmas, areas associated with cognitive control were found to be active, suggesting that individuals had to overcome an initial emotional evaluation in personal dilemmas in order to make a utilitarian decision in such cases. Greene, et al. therefore subsequently proposed a dual process theory of moral judgements in which certain dilemmas recruit immediate emotional reactions which can be in conflict with moral reflective "rational" analysis, and might require extra cognitive resources in order to arrive at a utilitarian decision. Further neuroscientific studies, using fMRI, have elaborated on these initial findings, generally supporting the view that certain features in moral dilemmas can trigger different brain processing areas. We understand that all human processes have their basis in the brain and therefore show correlations to brain activation patterns. As such, the new idea that the interfering effects of drugs on brain activity could manipulate not only basic brain functions but also have profound effects on higher order human processes such as moral decision-making seems only logical.

Drug effects are largely produced by interferences with neuronal transmission. Neurons use neurotransmitters to transfer information from one neuron to the next. At the synaptic cleft, the electric potential is transferred into a chemical signal by triggering the release of neurotransmitters into the synaptic cleft. The neurotransmitter then docks onto receptors of the post-synaptic cell membrane, triggering further intercellular mechanisms in the post-synaptic neuron, which enables the signal to be transferred.[4] Whilst the drug may interfere with multiple sites, one common mechanism is the blocking of receptors on the post-synaptic cell membrane which subsequently blocks signal transfer. If it is expected that certain human behaviours are mediated by activity of a neurotransmitter to a great extent, then blocking the activity of this neurotransmitter with a pharmaceutical would reduce such "behaviour".

In our previous study, we had the hypothesis that one key neurotransmitter, namely noradrenaline (NA), might be involved in moral judgement and moral behaviour. NA has previously been suggested to be involved in basic emotion processing and is thought to underlie the fight-or flight response. In fact, early research has already determined that fear responses in animals were associated with elevated levels of NA.[5] NA is transferred via alpha and beta receptors, peripherally and centrally. It has been suggested that emotional arousal should be reduced if NA receptors are blocked with a pharmaceutical.[6] Indeed, beta-blockers (i.e., propranolol, which blocks beta 1 and 2 receptors) have been found to reduce effects of emotional arousal. For instance, reduced heart rate and general reduced activation can be observed after emotional stimuli with propranolol intervention.[7] Therefore, propranolol has been prescribed not only as a first-line treatment for hypertension, but also as a means of reducing the effects of panic and anxiety. For example, propranolol is often prescribed for instances of performance anxiety, and for the prevention of the development of post-traumatic stress disorder (PTSD).

Crucially, if basic emotional arousal is also involved in higher order processes, such as our moral decision-making, then propranolol may also affect human morality.[7] And this is

4 For a description of the mechanism of chemical neurotransmission see Terbeck, et al., 'Noradrenaline Effects on Social Behaviour, Intergroup Relations, and Moral Decisions', 54–60.

5 A. F. Ax, 'The Psychological Differentiation Between Fear and Anger in Humans', *Psychosom Med.* 15:5 (1953), 433–442.

6 S. R. Chamberlain, U. Mueller, A. D. Blackwell, T. W. Robbins, and B. J. Sahakian, 'Noradrenergic Modulation of Working Memory and Emotional Memory in Humans', *Psychopharmacology* 188:4 (2006), 397–407.

7 Greene, et al., 'An fMRI Investigation of Emotional Engagement in Moral Judgment', 2105–2108.

indeed what we found in a recent study. Participants either received a single oral dose of propranolol or a placebo tablet. When the drug action had reached its peak effect, the psychological tests were conducted. Participants' heart rates were measured and a mood assessment was completed, followed by the moral dilemma test. In the morality task, participants judged the moral acceptability of a set of 20 moral dilemmas comprising both personal and impersonal scenarios including the footbridge and the switch dilemma. As predicted, we found that propranolol significantly reduced heart rate after the intervention but had no effect on self-reported mood. Importantly, participants in the propranolol group also judged harmful actions described in personal moral dilemmas as less morally acceptable, without changing the rating on impersonal moral dilemmas.[8] This suggests that NA function is involved in the psychological process of moral decision making and that responses to personal moral dilemmas may rely on basic emotional processes. However, contrary to what would be predicted according to Greene's theory, we found an increase in deontological or non-utilitarian judgements. We argue that, since NA is also involved in the processing of aggression, propranolol may have reduced aggression and subsequently increased harm aversion.

This theory regarding harm aversion and its role in moral decision making has also been supported by further psychopharmacological research using the drug citalopram. Citalopram is a selective serotonin re-uptake inhibitor and increases the net brain concentration of serotonin. In previous studies, serotonin has been found to be involved in emotion regulation, and is thus also often used for the treatment of mood or anxiety disorders.[9] In this study, Crockett, Clark, Hauser, and Robbins gave participants a single dose of citalopram in a double-blind placebo controlled study before assessing each participant's moral judgements.[10] Again, the authors found an effect of the drug on moral judgements, but only for personal dilemmas. Specifically, they found that citalopram increased deontological responses in these dilemmas, arguing that serotonin reuptake inhibition may have increased harm aversion.

Apart from psychopharmacological manipulation of theoretical moral judgements, studies have also investigated wider moral behaviour and social attitudes, such as pro-social behaviour, generosity, and fairness judgements, as well as perceptions of out-group members.[11] In numerous studies it has been found that pharmaceuticals produce effects on such human behaviours.[12] For instance, oxytocin, a hormone associated with maternal care and bonding has been found to also increase fairness judgements, willingness to donate, and to help others[13]. In another recent study, we investigated the effect of propranolol on intergroup attitudes, using behavioural as

8 S. Terbeck, G. Kahane, S. McTavish, J. Savulescu, N. Levy, M. Hewstone, and P. J. Cowen, 'Beta Adrenergic Blockade Reduces Utilitarian Judgement', *Biological Psychology* 92:2 (2013), 323–328.

9 For example see I. Kirsch, B. J. Deacon, T. B. Huedo-Medina, A. Scoboria, T. J. Moore, and B. T. Johnson, 'Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food
and Drug Administration', *PLoS Med*. 5:2 (2008), 260–268.

10 M. J. Crockett, L. Clark, M. D. Hauser, and T. W. Robbins, 'Serotonin Selectively Influences Moral Judgment and Behavior Through Effects on Harm Aversion', *Proc Natl Acad Sci USA* 107:40 (2010), 17433–17438.

11 C. K. W. De Dreu, L. L. Greer, G. A. Van Kleef, S. Shalvi, and M.
J. J. Handgraaf, 'Oxytocin Promotes Human Ethnocentrism', *Proc Natl Acad Sci USA* 108:4 (2011), 1262–1266.

12 For a review see N. Levy, T. Douglas, G. Kahane, S. Terbeck, P. J. Cowen, M. Hewstone, and J. Savulescu, 'Are You Morally Modified? The Moral Effects of Widely Used Pharmaceuticals', *Philos Psychiatr Psychol.* 21:2 (2014), 111–125.

13 For example see P. J. Zak, A. A. Stanton, and S. Ahmadi, 'Oxytocin Increases Generosity in Humans', *PLoS One* 2:11 (2007), 1–5.

well as combined fMRI methods.[14] We first investigated racial biases using self-reporting as well as response-time based computer test methods (i.e., the IAT). The IAT measures racial biases by comparing response times between associations of in-and-out-group faces as well as positive and negative words. We found that propranolol reduced racial biases, suggesting that NA might be involved in the processing of social attitudes.[14] More recently we repeated this study, but in addition also examined the effect of propranolol on racial face perception in the brain using fMRI.[14] In this study, Caucasian participants received the pharmacological intervention before undergoing fMRI in which they viewed black and white faces. We found activation differences in the fusiform gyrus (a brain area strongly associated with face perception and social categorisation) with propranolol. This suggests that NA might be involved in basic face processing and immediate social categorisation in the brain. The above studies have illustrated that, besides having an effect on basic physiological processes, drugs can also influence higher order human social processes, such as theoretical moral judgements, judgements of fairness and generosity, as well as pro-social behaviour and social perception. Does this therefore mean that we can take drugs to enhance our morality?

2. Psychopharmacological Effects "Outside" the Laboratory

In neuroscience, when conducting experiments using pharmaceuticals, the idea is to investigate and learn more about the underlying neural mechanisms of certain concepts. For instance, as determined in our own research, we found that NA seems to play a significant and causal role in moral decision-making and social judgement. As such, these studies are theoretically driven. Attempts to *find* a drug which might be used to *enhance* our morality is less theoretical in nature; indeed, to our understanding, previous studies were not designed to test this latter idea, but rather to understand the neuroscience of higher order human processes. Indeed, whilst conducting these studies, we did not consider the concept of moral enhancement or if it was even possible.

When conducting experimental research, the results reported are almost always average effects. For instance, even though we found that propranolol significantly reduced racial biases on the IAT test, this was only the case for the average of the group, meaning that racial biases were not reduced in every single individual. Secondly, several studies must be carried out before meaningful implications can be inferred. For instance, in a recent review we investigated the potential use of metabotropic glutamate receptor 5 antagonists for the treatment of anxiety.[15] In this review we described that, before its application in humans, more than 800 animal studies were conducted. Indeed, in order to assume reliable pharmacological effects, a large number of experiments are required.

It was also suggested that selective serotonin reuptake inhibitors (SSRIs) might be used to elevate or enhance mood in humans.[16] In comparison to the limited number of pharmacological

[14] S. Terbeck, G. Kahane, S. McTavish, R. McCutcheon, M. Hewstone, J. Savulescu, and R. Norbury, 'Beta-Adrenoceptor Blockade
Modulates Fusiform Gyrus Activity to Black versus White Faces', *Psychopharmacology (Berl.)* 232:16 (2015), 2951–2958. See also S. Terbeck, G. Kahane, S. McTavish, J. Savulescu, P. J. Cowen, and M. Hewstone, 'Propranolol Reduces Implicit Negative Racial Bias', *Psychopharmacology* 222:3 (2012), 419–424.

[15] S. Terbeck, F. Akkus, L. P. Chesterman, and G. Hasler, 'The Role of Metabotropic Glutamate Receptor 5 in the Pathogenesis of Mood Disorders and Addiction: Combining Preclinical Evidence with Human Positron Emission Tomography (PET) Studies', *Front Neurosci.* 9:86 (2015), 1–10.

[16] R. De Jongh, I. Bolt, M. Schermer, and B. Olivier, 'Botox for the Brain: Enhancement of Cognition, Mood and Pro-Social Behavior and Blunting of Unwanted Memories', *Neuroscience & Biobehavioral Reviews* 32:4 (2008), 760–776.

studies on morality, numerous studies have been conducted on "mood enhancement". A recent review examined the overall effect of SSRIs on mood.[17] The authors found that SSRIs were only effective in changing mood in severe depression but not in moderate or mild depression, suggesting that mood enhancement is selective and therefore not possible across individuals. Furthermore, it is often assumed that the effects of laboratory experiments can translate to the outside world, which may prove problematic. In particular, with regards to drug effects, the short and long term effects may vary. For instance, with regards to aggression reduction in association with propranolol, it was speculated that the effects of aggression reduction would disappear over time, whilst the 'medical' effect of reduced heart rate remained.[18]

In addition, we would like to emphasise another factor associated with pharmacological intervention, which is the inevitability of side-effects. Previously, in ethics and philosophy, some articles discussing the prospect of pharmacological enhancement suggested that in the near future there may be a drug with no or negligible side effects.[19] Other claims, such as the argument that taking a drug might not be different from other non-medical interventions might indeed be defensible, if the assumption that there will be a drug with no or negligible side effects were justified. For example, if we found a drug that could make you fly and nothing else, we would most likely take it. But it is not that straightforward. In fact, in our recent article we discussed why there is not currently, and likely will never be, a drug with no or negligible side effects.[20] Side effects of most psychoactive substances range from allergic reactions to physical problems, but also to psychological side-effects including reduced sexual drive, increased anxiety, loss of attention, increased tiredness, or loss of motivation. From our current understanding of the brain, we know that interventions that do simply *one* thing and nothing else are not possible. The brain operates in complex networks. We cannot find one area for morality or one neurotransmitter regulating morality, or one drug that has only one effect on this complex network. In our recent article, we illustrated a case of simple visual edge detection in mice and described the potential side effects that 'enhancing' this seemingly simple function might elicit.[21] The complexity demonstrated here with a process as simple as edge detection only emphasises the challenges of attempting to enhance a multifaceted phenomenon such as morality. Therefore, decisions regarding enhancement must be pragmatic evaluations of effect versus side effect. Crucially, in this pursuit of moral enhancement, we first need reliable measures of morality to determine whether there has been an effect, and this measure would need to prove meaningful outside of the laboratory. But do we have such reliable measures of morality?

3. The Psychological Assessment of Morality and its Reliability

In order to examine measures of morality, we must first look at the history of assessment within the moral domain. Until recently, morality had been a topic of speculation in

17 I. Kirsch, B. J. Deacon, T. B. Huedo-Medina, A. Scoboria, T. J. Moore, and B. T. Johnson, 'Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration', *PLoS Med*. 5:2 (2008), 260–268.

18 J. M. Silver, S. C. Yudofsky, J. A. Slater, R. K. Gold, B. L. Stryer, D. T. Williams, H. Wolland, and J. Endicott, 'Propranolol Treatment of Chronically Hospitalized Aggressive Patients', *J. Neuropsychiatry Clin Neurosci*. 11:3 (1999), 328–335.

19 For example, see T. Douglas, 'Moral Enhancement', in J. Savulescu, R. ter Meulen, and G. Kahane (eds), *Enhancing Human Capacities* (Oxford, UK: Wiley-Blackwell, 2011), 467–485.

20 S. Terbeck and L. P. Chesterman, 'Will There Ever Be a Drug With No or Negligible Side Effects? Evidence From Neuroscience', *Neuroethics* 7:2 (2014), 189–194.

21 Terbeck, et al., 'Will There Ever Be a Drug With No or Negligible Side Effects? Evidence From Neuroscience', 189–194

philosophy, grounded in theory rather than empirical investigation. It was the emergence of moral psychology that marked the fusion of both theoretical and experimental approaches in investigating the nature of morality, but more specifically, moral judgement.[22]

In line with the hybrid nature of moral psychology and as discussed above, provocative moral dilemmas pitting deontological against utilitarian theories have played a central role in the investigation of moral judgement.[23] Both the footbridge and switch dilemmas,[24] which are versions of the so-called "trolley problem", have become a topic of interest for both moral philosophers and moral psychologists for the reason that individuals tend to endorse the utilitarian outcome in the switch dilemma but refuse to do so in the footbridge alternative.[25]

At the methodological level, moral psychologists tend to present these trolley problems to participants in text-based paradigms that require a subsequent moral judgement. Typically, participants are asked whether the utilitarian act described in the dilemma is "appropriate" or "acceptable". In their conception, these hypothetical moral dilemmas were not intended to reveal insights into real-life decisions but instead, in their experimental simplicity, allow moral scientists to explore the 'foundational psychological processes that underlie human moral cognition'.[26] The level of experimental control available to scientists in incorporating these paradigms is paramount; allowing moral conflicts to arise in artificial contexts with anonymous agents.[27] Despite the prevalent use of these paradigms, concerns have been raised about the precision with which moral dilemmas have been constructed.[28] Differences in several factors including framing, word count, perspective, situational circumstances, and type of question have been shown to influence moral judgements.[29] For example, research has distinguished judgement questions such as 'is it morally acceptable?' from action-choice questions such as 'would you do it?' Arguably, judgement questions address allocentric evaluations of the utilitarian act, whereas action-choice questions offer an egocentric perspective, resulting in distinct moral judgements.[30] In a well-known framing study, participants were found to endorse utilitarian outcomes when the phrasing "save" was used as opposed to "kill".[31]

22 J. M. Doris, 'Introduction', in J. M. Doris (ed.), *The Moral Psychology Handbook* (New York, USA: Oxford University Press, 2010), 1–2.

23 See D. M. Bartels, C. W. Bauman, F. A. Cushman, D. A. Pizarro, and A. P. McGraw, 'Moral Judgment and Decision Making', in G. Keren and G. Wu (eds), *The Wiley Blackwell Handbook of Judgment and Decision Making* (Chichester, UK: John Wiley & Sons, 2015), 478–515. See also F. Cushman, L. Young, and J. Greene, 'Our Multi-System Moral Psychology: Towards a Consensus View', in J. M. Doris (ed.), *The Moral Psychology Handbook* (New York, USA: Oxford University Press, 2010), 47–71.

24 Thomson, *Killing, Letting Die, and the Trolley Problem*, 204–217. See also Foot, *Virtues and Vices*.

25 Greene, et al., 'An fMRI Investigation of Emotional Engagement in Moral Judgment', 2105–2108.

26 J. F. Christensen, and A. Gomila, 'Moral Dilemmas in Cognitive Neuroscience of Moral Decision-Making: A Principled Review', *Neurosci Biobehav Rev.* 36:4 (2012), 1250.

27 M. Hauser, F. Cushman, L. Young, R. Kang-Xing Jin, and J. Mikhail, 'A Dissociation Between Moral Judgments and Justifications', *Mind & Language* 22:1 (2007), 1–21.

28 Christensen and Gomila, 'Moral Dilemmas in Cognitive Neuroscience of Moral Decision-Making: A Principled Review', 1249–1264.

29 Bartels, et al., 'Moral Judgment and Decision Making', 154–161.

30 S. Tassy, O. Oullier, J. Mancini, and B. Wicker, 'Discrepancies Between Judgment and Choice of Action in Moral Dilemmas', *Frontiers in Psychology* 4:250 (2013), 1–8.

31 L. Petrinovich and P. O'Neill, 'Influence of Wording and Framing Effects on Moral Intuitions', *Ethology and Sociobiology* 17:3 (1996), 145–171.

In addition to disagreement over the formulations of these moral dilemmas, research has also questioned the ability of these paradigms to reflect genuine deontological or utilitarian responses.[32] For example, in the footbridge dilemma we have no clear method for distinguishing the action of pushing the man as (i) a moral decision grounded in the belief that killing one to save the majority is morally required, or as (ii) a decision driven simply by less aversion to harm.[33] Crucially, there is evidence to suggest that people possessing antisocial personality traits are more likely to push the man off the footbridge with their intentions masked as being utilitarian.[34] Further, research has found that even when controlling for this anti-social trait association, so-called utilitarian responses do not always reflect concern for the 'greater good'.[35]

Despite these criticisms, recent attempts to validate moral dilemmas have addressed previous inconsistencies in formulation[36] and previous attempts to label individuals as either utilitarian or deontological have been challenged with research arguing that people instead tend to adopt a 'particularist' approach to morals that takes the details of each case into account'.[37] Whilst the methodology is valuable in shedding light on the mechanisms underlying moral judgements, there remain gaps in our understanding. How can declarations made in response to these text-based paradigms translate into real-world moral behaviour?[38]

Attempts to explicate moral behaviour have largely explored non-harmful actions in economical paradigms and so the investigation of harmful moral actions has made little headway.[39] While text-based moral dilemmas possess an advantage in producing unambiguous outputs, these questionnaire-based paradigms 'only offer a very low degree of immersion'.[40] In fact, FeldmanHall, et al. found that these contextually impoverished scenarios elicited moral decisions that were different from those made in real counterparts of the same scenario.[41] Critically, by increasing the amount of contextual information available in the

32 G. Kahane, J. A. Everett, B. D. Earp, M. Farias, and P. Savulescu, 'Utilitarian Judgments in Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good', *Cognition* 134 (2015), 193–209.

33 For example, see I. Patil, 'Trait Psychopathy and Utilitarian Moral Judgement: The Mediating Role of Action Aversion', *Journal of Cognitive Psychology* 27:3 (2015), 349–366. See also F. Cushman, K. Gray, A. Gaffey, and W. B. Mendes, 'Simulating Murder: The Aversion to Harmful Action', *Emotion* 12:1 (2012), 2–7.

34 See D. M. Bartels and D. A. Pizarro, 'The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas', *Cognition* 121:1 (2011), 154–161. See also H. Djeriouat and
B. Tremoliere, 'The Dark Triad of Personality and Utilitarian Moral Judgment: The Mediating Role of Honesty/Humility and Harm/Care', *Personality and Individual Differences* 67 (2014), 11–16. See also Y. Gao and S. Tang, 'Psychopathic Personality and Utilitarian Moral Judgment in College Students', *Journal of Criminal Justice* 41:5 (2013), 342–349.

35 Kahane, et al., 'Utilitarian Judgments in Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good', 12.

36 J. F. Christensen, A. Flexas, M. Calabrese, N. K. Gut, and A. Gomila, 'Moral Judgment Reloaded: A Moral Dilemma Validation Study', *Frontiers in Psychology*' 5:607 (2014), 1–18.

37 Christensen, et al., 'Moral Judgment Reloaded', 16.

38 T. D. Parsons, 'Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences', *Front Hum Neurosci.* 9:660 (2015), 1–14.

39 C. D. Navarrete, M. M. McDonald, M. L. Mott, and B. Asher, 'Virtual Morality: Emotion and Action in a Simulated Three-Dimensional "Trolley Problem"', *Emotion* 12:2 (2012), 364–370.

40 A. Skulmowski, A. Bunge, K. Kaspar, and G. Pipa, 'Forced-Choice Decision-Making in Modified Trolley Dilemma Situations: A Virtual Reality and Eye Tracking Study, *Front Behav Neurosci.* 8:426 (2014), 2.

41 O. FeldmanHall, D. Mobbs, D. Evans, L. Hiscox, L. Navrady, and
T. Dalgleish, 'What We Say and What We Do: The Relationship Between Real and Hypothetical Moral Choices', *Cognition* 123:3 (2012), 434–441.

hypothetical scenario, the researchers were able to align hypothetical moral choices with real moral choices. This line of research would suggest that contextual richness alters moral decisions and raises further questions regarding the reliance on text-based moral dilemmas in moral psychology. Echoing the age-old saying "do as I say, not as I do",[42] this inconsistency between moral judgement and action is supported in research with institutionalised psychopaths who display intact understanding of moral norms but exhibit antisocial and sometimes violent behaviours.[43]

Taken together, this evidence of a partial dissociation between moral judgements and moral actions and the potential for contextual information to bridge the gap highlights the need for a contextually rich testing tool. Fortunately, the emergence of contextually salient virtual reality technologies has opened opportunities to explore simulated harmful moral actions in environments free from issues concerning de-contextualisation.[44] Virtual reality systems adopt sensory-tracking, most commonly head-tracking, to immerse participants within life-size simulated environments. In these dynamic environments, researchers can begin to investigate active moral choices: 'would someone [...] actually resort to this course of action when the full repertoire of contextual features comes into play?'[45]

In this virtual domain of moral psychology, attempts to reproduce moral dilemmas in virtual reality have revealed mixed findings regarding the relationship between moral judgement and action. While in virtually constructed versions of trolley-like dilemmas some research has demonstrated consistency between judgements in original text-based paradigms and simulated actions in virtual counterparts,[46] contrasting research has demonstrated a disparity with greater utilitarian endorsements observed in virtual dilemmas.[47] For example, in a recent study, we found that when individuals were required to simulate the harmful action of pushing the man off the bridge in a virtual simulation of the personal footbridge dilemma, the majority of people chose the utilitarian action.[48] When faced with the text-based version of the dilemma,

42 For an overview of moral hypocrisy see B. Monin and A. Merritt, 'Moral Hypocrisy, Moral Inconsistency, and the Struggle for Moral Integrity', in M. Mikulincer and P. R. Shaver (eds), *The Social Psychology of Morality: Exploring the Causes of Good and Evil: Herzliya Series on Personality and Social Psychology* (Washington, DC: American Psychological Association, 2012), 167–184.

43 M. Cima, F. Tonnaer, and M. D. Hauser, 'Psychopaths Know Right From Wrong But Don't Care', *Soc Cogn Affect Neurosci*. 5:1 (2010), 59–67. See also K. A. Kiehl, 'Without Morals: The Cognitive Neuroscience of Criminal Psychopaths', in W. Sinnott-Armstrong (ed.), *Moral Psychology* (Cambridge, USA: The MIT Press, 2008), 119–149. See also S. Tassy, C. Deruelle, J. Mancini, S. Leistedt, and B. Wicker, 'High Levels of Psychopathic Traits Alters Moral Choice But Not Moral Judgment', *Front Hum Neurosci.* 7 (2013): https://doi.org/10.3389/fnhum.2013. 00229.

44 For an example see K. B. Francis, C. Howard, I. S. Howard, M. Gummerum, G. Ganis, G. Anderson, and S. Terbeck, 'Virtual Morality: Transitioning from Moral Judgment to Moral Action?', *PLoS One* 11:10 (2016), 1–22.

45 I. Patil, C. Cogoni, N. Zangrando, L. Chittaro, and G. Silani, 'Affective Basis of Judgment-Behavior Discrepancy in Virtual Experiences of Moral Dilemmas', *Soc Neurosci.* 9:1 (2014), 95.

46 See Navarrete, et al., 'Virtual Morality: Emotion and Action in a Simulated Three-Dimensional "Trolley Problem"', 364–370. See also Skulmowski, et al., 'Forced-Choice Decision-Making in Modified Trolley Dilemma Situations: A Virtual Reality and Eye Tracking Study', 1–16.

47 See Francis, et al., 'Virtual Morality: Transitioning from Moral Judgment to Moral Action', 1–22. See also Patil, et al., 'Affective Basis of Judgment-Behavior Discrepancy in Virtual Experiences of Moral Dilemmas', 94–107. See also X. Pan and M. Slater, 'Confronting a Moral Dilemma in Virtual Reality: A Pilot Study', *Proceedings of the 25th BCS Conference on Human-Computer Interaction* (2011).

48 For a full description see Francis, et al., 'Virtual Morality: Transitioning from Moral Judgment to Moral Action', 1–22.

on the other hand, the majority of people refused to endorse the utilitarian outcome. Arguably, the degree of contextual saliency and subsequent affective responses experienced in virtual moral dilemmas produces this discrepancy.

Virtual reality systems offer considerable advantages. Unlike other research domains in which actions can be examined both in the laboratory and in the field, the domain of morality presents unique challenges; participants cannot be placed in real-world precarious situations.[49] While economic paradigms have begun to shed light on non-harmful actions, studies investigating morality of harm have remained largely non-behavioural for the reason that ethically harmful behaviours prove difficult to test.[50] Although the incorporation of harm-based moral dilemmas, whether text-based or virtual, is a somewhat limited approach, helping us to investigate 'only a fragment of our moral psychology', it is a 'potentially significant one'.[51] The application of virtual reality to study harm-specific moral actions for example has significance for professions exposed to sensitive and emotionally arousing moral decision-making on a regular basis.[52] Emergency service professionals have begun to adopt virtual reality systems for several reasons. Not only does it offer full immersion, but it is also both cost-effective and safe in contexts where there is little room for error.[53] This application of virtual reality demonstrates its broader value, investigating real-world moral decision-making beyond that of hypothetical scenarios centred round normative theories.

The level of "realism" available in virtual environments not only relies on visual saliency but also on "sensorimotor contingencies" or the congruence between motor actions and sensory simulation. Presently, virtual reality systems can only 'offer crude approximations' of sensorimotor contingencies and subsequent plausibility.[54] Despite this shortcoming, research has shown that even basic virtual environments can elicit a range of realistic responses providing opportunities to bridge the "reality gap" in social domains. Essentially, if the virtual environment can deliver the subjective experience of "being there", life-like thoughts and emotions can be prompted.[55] Importantly, research demonstrates that virtual reality systems can offer successful collaboration between the experimental control available in laboratory settings and components of ecological validity in providing enhanced affective experiences.[56] In fact, with regards to making virtual environments true to life, we face a paradox in research settings. Preserving the distinction between reality and virtual reality is essential for ethical reasons.[57] If the boundary were to break down, then the potentially hazardous reasons for not evaluating moral choices in the field in the first place would become of concern in virtual reality paradigms.[58] Despite the compromise here between bridging the reality gap and

49 M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, and M. V. Sanchez-Vives, 'A Virtual Reprise of the Stanley Milgram Obedience Experiments', *PLoS One* 1:1 (2006), 1–10.

50 A. Rovira, D. Swapp, B. Spanlang, and M. Slater, 'The Use of Virtual Reality in the Study of People's Responses to Violent Incidents',
Front Behav Neurosc. 3:59 (2009), 1–10.

51 M. Hauser, et al., 'A Dissociation Between Moral Judgments and Justifications', 4.

52 Francis, et al., 'Virtual Morality: Transitioning from Moral Judgment to Moral Action', 1–22.

53 P. B. Andreatta, E. Maslowski, S. Petty, W. Shim, M. Marsh, T. Hall, and J. Frankel, 'Virtual Reality Triage Training Provides a Viable Solution for Disaster-Preparedness', *Academic Emergency Medicine* 17:8 (2010), 870–876.

54 Rovira, et al., 'The Use of Virtual Reality in the Study of People's Responses to Violent Incidents', 2, 9.

55 A. Carassa, F. Morganti, and M. Tirassa, 'A Situated Cognition Perspective on Presence', *Proceedings of the Cognitive Science Society* 27 (2005), 384.

56 See Parsons, 'Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social
Neurosciences', 1–14.

57 Slater, et al., 'A Virtual Reprise of the Stanley Milgram Obedience Experiments', 1–10.

preserving the technological boundary, virtual reality systems can generate experiences and trigger emotions that decontextualised and impoverished text-based paradigms cannot.[59]

To date, existing models of moral decision-making have been shaped with moral judgements in mind,[60] offering little insight into moral actions. Although virtual research is in its infancy, by promoting "judicious use" of these virtual reality paradigms within moral psychology, we can begin to validate their potential in assessing morality. Crucially, the finding that moral judgements in text-based paradigms diverge from moral actions in virtual reality paradigms raises two key questions: (i) are moral judgement and moral action driven by at least partially distinct mechanisms and (ii) do immersive virtual environments reveal moral choices that are closer to real-life responses? In either stream, the incorporation and validation of both methods must continue to better our assessment of both moral judgements and actions; after all 'by examining only one blade of a pair of scissors, one will not understand how scissors cut'.[61] Only through adopting these multifaceted approaches to the study of morality can we begin to assess moral decision-making and, subsequently, begin to examine, understand, and assess the impact of moral enhancement.

4. Conclusion

In this chapter we have demonstrated that psychopharmacological manipulation can in fact alter moral judgements[62] as well as social behaviours and attitudes.[63] In this sense, "morality" does appear susceptible to alteration and enhancement. However, many of the studies that we review were originally designed to unearth the underlying neural mechanisms responsible for such moral judgements and social behaviours, placing little focus on the concept of moral enhancement and short and long-term effects of drug use outside of the lab. When considering moral enhancement in this practical and pragmatic framework, we highlight two key areas in need of advancement. Firstly, there will likely never be a drug without side-effects.[64] As such, without first establishing control over or abolishing such harmful side-effects, we cannot yet advocate psychopharmacological-based moral enhancement. Secondly, given the multifaceted nature of morality as a construct and debates concerning the reliability of its measurement,[65] we must first validate and refine its assessment prior to fully understanding the nature and extent of any such moral enhancement.

58 Rovira, et al., 'The Use of Virtual Reality in the Study of People's Responses to Violent Incidents', 1–10.

59 Francis, et al., 'Virtual Morality: Transitioning from Moral Judgment to Moral Action', 1–22.

60 For an example see Greene, et al., 'An fMRI Investigation of Emotional Engagement in Moral Judgment', 2105–2108.
See also F. Cushman, 'Action, Outcome, and Value: A Dual-System Framework for Morality', *Pers Soc Psychol Rev.* 17:3 (2013), 273–292.
See also J. Moll, R. de Oliveira-Souza, and P. J. Eslinger, 'Morals and the Human Brain: A Working Model', *Neuroreport* 14:3 (2003), 299–305.

61 A. Bunge and A. Skulmowski, 'Descriptive and Pragmatic Levels of Empirical Ethics: Utilizing the Situated Character of Moral Concepts,
Judgment, and Decision-Making', in C. Luetge, H. Rusch, and M. Uhl (eds), *Experimental Ethics: Toward an Empirical Moral Philosophy* (Basingstoke, UK: Palgrave Macmillan, 2014), 176.

62 Terbeck, et al., 'Beta Adrenergic Blockade Reduces Utilitarian Judgement', 323–63 Terbeck, et al., 'Beta-Adrenoceptor Blockade Modulates Fusiform Gyrus Activity to Black versus White Faces', 2951–2958.

64 Terbeck and Chesterman, 'Will There Ever Be a Drug With No or Negligible Side Effects? Evidence From Neuroscience', 189–194.

65 Francis, et al., 'Virtual Morality: Transitioning from Moral Judgment to Moral Action', 1–22